Supplementary Information for *Microbial island biogeography: isolation shapes the life history characteristics but not diversity of root-symbiotic fungal communities*

Supplementary Methods

Brief geological history of the study islands.

Iceland, oceanic, 101826 km2, 980 km from continent (Eurasia; 290 Greenland). Emergence around 18 million years ago; oldest continuous ice-free area appeared around 10 Ka ago.

Iceland lies on the Mid-Atlantic Ridge, the divergent boundary between the Eurasian and the North American plates. It also lies above a hotspot, the Iceland plume, which is believed to have caused the formation of Iceland itself around 16-18 million years ago. The island is geologically characterized by volcanic rock and volcanoclastic sediments. During the last 5 million years glaciers have developed on the island, with the island completely covered by ice several times in the last 2 million years. The oldest continuously ice-free area appeared 10 Ka ago (Saemundsson, 1979).

Svalbard, continental, 37814 km2, 660 form continent (Eurasia). Probably most recent emergence in Cretaceous ca 130 million years ago; oldest continuous ice-free area appeared about 14 Ka ago.

The island has a diverse geology, with Precambrian-Lower Palaeozoic metamorphic basement, Devonian-Paleogene bedrock (sandstones, shales, evaporates) and Quaternary sediments (including glacial and marine sediments). During the last 3 million years the island has been repeatedly covered by glaciers. The last glaciation covering the entire landmass retreated 14-10 Ka ago, leaving local glaciers in higher areas and valleys (Dallmann, 2015).

Crete, continental, 8350 km2**,** 100 from continent (Eurasia). Most recent emergence 2 million years ago.

Crete is part of the Hellenic Island Arc, which formed during the closure of the Tethys Ocean. The area was formed as part of the continent around 25 million years ago. Around 10 million years ago, the Mediterranean Sea flooded the Aegean area and the island became separate. In the following period, Crete was repeatedly connected to the land (e.g., Messinian Salinity Crisis, 5.9-5.3 Ma ago) or disconnected and flooded by sea level fluctuations. Crete was completely submerged during the Pliocene, and gradually emerged following tectonic uplift in the Early Pleistocene, about 2 million years ago (van der Geer *et al.,* 2006).

Mallorca**,** continental, 3667 km2, 170 km from continent (Eurasia). Part of Iberian microplate since about 250 million years ago; separated from mainland 5 million years ago.

Mallorca has a Variscian (Late Palaeozoic) metamorphic continental basement and Meso-Cenozoic sedimentary cover (including carbonates). The Balearic Islands represent the exposed peaks of a submarine promontory that is an elongation of the Spanish Betic system. The island area was under tectonic deformations during the Neogene, 25-15 million years ago. The Balearic islands were connected to the continent during the Messinian Salinity Crisis. The following transgression 5.3 million years ago disconnected the island from the mainland and the terrestrial environment of each island has developed separately since that time (Bover *et al.,* 2008).

Öland**,** continental, 1342 km2**,** 6 km from continent (Eurasia). Emerged 12 Ka ago.

The exposed bedrock on Öland consists of Ordovician carbonate rocks. The last continental glacier retreated from the area of Öland in southern Sweden about 15 Ka ago, after which the area was covered by the Baltic Ice Lake. The highest part of the island (now 55 m above sea level) emerged from the water 12.2 Ka ago (Vassiljev and Saarse, 2013).

Saaremaa, continental, 2673 km2, 15 km from continent (Eurasia). Emerged 11.6 Ka ago.

The exposed bedrock on Saaremaa consists of Silurian carbonate rocks. The last continental glacier retreated from the area of Saaremaa about 14 Ka ago, after which the area was covered by Baltic Ice Lake. The highest part of the island (now 53 m above sea level) emerged from the water 11.6 Ka ago, which was caused by the Billingen drainage of the Baltic Ice Lake (Rosentau *et al.,* 2009).

Sal, Cape Verde, oceanic, 216 km2**,** 610 km from continent (Africa). Emerged as a volcano approximately 20 million years ago.

The Cape Verde islands are situated on a geological hotspot and have developed on the deep mantle plume. All of the islands are of volcanic origin, composed of igneous rocks of basic composition. The oldest Cretaceous igneous rocks are covered by Meso-Cenozoic marine sediments (carbonates). The first phase of volcanism began in the Early Miocene (around 20-26 million years ago), reaching its peak activity in the Middle to Late Miocene (7–15 million years ago) when the islands attained their maximum size. Sal represents the old, razed edifices in the eastern part of archipelago (Ramalho, 2011).

Santiago, Cape Verde, oceanic, 991 km2, 640 from continent (Africa). Emerged as a volcano approximately 10 million years ago.

The oldest rocks on Santiago are 130 million years old pillow lava (submarine). Santiago represents the younger shield volcanoes in the west of the archipelago. Around 10–8 million years ago, volcanism seems to have started west of the original area, creating the basement complexes of Santiago (Ramalho, 2011).

Guadeloupe, oceanic, 1438 km2, 580 from continent (South America). Basse-Terre subaerial volcanism since ca 3 million years ago; Grande-Terre emerged from shallow carbonate sea 0.6-1 million years ago.

Guadeloupe is part of the Lesser Antillean volcanic island arc. Most of the magmatic products found in the island arc are related to the subduction of the North American Plate

below the Caribbean plate at least 40 million years ago. Around the Mid-Miocene (around 10 million years ago) the subducting slab and volcanic front migrated west from their previous position (including Grande-Terre) and created a new volcanic arc (including Basse-Terre). Grande-Terre has an older Paleogene volcanic basement covered by a thick Pliocene-Pleistocene (1-5 million years ago) marine carbonate complex. The carbonate platform of Grande-Terre emerged in the late Calabrian (0.6-1 million years ago). Extensive volcanic activity on Basse-Terre started at least 3-3.5 million years ago ago (Münch *et al.,* 2013).

Tasmania, continental, 65022 km2, 200 km from continent (Australia). Separated from Antarctica 45 million years ago ago.

Tasmania is part of the Australian, Antarctican and Indian (eastern Gondwana) continent, having a continental basement of up to 1270 million years old. In the Permian period (270-300 million years ago) the area was covered by glaciers. Later, in the Triassic (200-250 million years ago) the Tasmania marine basin formed. A giant intrusion of magma occurred in the Jurassic, 183 million years ago, forming diabase (dolerite) which covers large areas of Tasmania today. Continental breakup occurred during the Cretaceous and Cenozoic Periods, splitting off undersea plateaus, forming the Bass Strait and ultimately separating Tasmania from Antarctica (finally about 45 million years ago). Rifting started 83 million years ago from the east, separating Tasmania from Australia. Flowering plants moved into Tasmania about 90 million years ago, when it was still connected to Antarctica (and Australia). The higher mountains were glaciated during the most recent Pleistocene glacial period (44-16 thousand years ago) (Hill, 1990).

New Caledonia, continental, 16648 km2, 1270 km from continent (Australia)**.** Emerged 37 million years ago; alternatively preserved refugium of pre-Eocene Australian biota.

New Caledonia separated from the Australian continent 65-80 million years ago, and drifted to the northeast, reaching its present position about 50 million years ago. However, based on some phylogenetic studies and geological evidence, New Caledonia would be better regarded as an oceanic island, dating to the late Eocene (about 37 million years ago). At this time nearly all of New Caledonia was covered with up to 2 km of peridotites, a type of

igneous rock formed in the ocean, meaning that all the island must have been below the ocean surface. Consequently, the island's entire biota can only have resulted from recolonization since the Oligocene, which is consistent with independent dating from molecular phylogenetic studies (Grandcolas *et al.,* 2008).

However an alternative view exists about the origin of biota in New Caledonia: Inference from the modern flora, indicates that at least some land must have remained exposed, serving as a refugium. Many attributes of New Caledonia's flora, such as its high generic and familial diversity, and the presence of numerous primitive groups, would be very difficult to explain otherwise (e.g., by long-distance dispersal). A substantial component of today's flora is thus thought to comprise the descendants of pre-Eocene Australasian groups that remained on New Caledonia as it separated and drifted away from Australia. The extensive ultramafic substrates (peridotites and serpentinites) still cover about one third of the total land area related to characteristic ultrabasic soils, which have exceptionally high levels of Fe, Mg, and of several heavy metals such as Ni, Cr, Co, and Mn, elements that are generally toxic to plants, along with very low levels of N, P, K, Ca, and Al. These chemical imbalances result in highly specialized edaphic conditions that have had a profound influence on the evolution and diversification of New Caledonia's flora and vegetation (Lowry II, 1998)

Tahiti, oceanic, 1042 km2, 5870 km from continent (Australia). Emerged as volcano approximately 1.2 million years ago.

The origin of most Polynesian islands are volcanic hot spots under the earth's crust. At a regular interval of several million years, the hotspot becomes active and forms a new volcano that pierces the surface of the ocean and becomes, once extinguished, a high island. The islands recede north-west by 10 cm a year leaving extremely deep trenches (around 4000 m) between them. The Society Islands are thought to have originated from a hot spot located near the island of Mehetia, 110 km east of Tahiti. The chain exhibits an age progression to the west from Mehetia (less than 75 000 years ago) to Tahiti (0.6–1.2 million years ago), Moorea (1.5–2.0 Ma), Raiatea (2.5–2.7 Ma), Bora Bora (3.1–3.5 Ma) and Maupiti (4.2–4.5 Ma), consistent with a plate motion of 110 mm yr−1 (Binard *et al.,* 1993; Blais *et al.,* 1997, 2002; Neall and Trewick, 2008).

Tahiti consists of two old volcanoes—Tahiti-Nui in the northwest and Tahiti-Iti in the southeast—linked by an isthmus. Tahiti-Nui formed as a volcanic shield between 1.4 and 0.9 million years ago. Both the northern and southern flanks of Tahiti-Nui collapsed sometime around 0.9 million years ago, gouging massive arcs out of the island's perimeter. Another volcano, Tahiti-Iti joined Tahiti-Nui by isthmus soon after their formation (Hildebrand *et al.,* 2004).

Tikehau, Tuamotu, oceanic, 20 km2**,** 5880 from continent (North America). Emerged as a volcano >50 million years ago; later subsided and developed into a coral atoll.

Tikehau is a low coral atoll in the western Tuamotu (French Polynesia), located 340 km northeast of Tahiti. The atoll's oval-shaped lagoon is 27 km long and 19 km wide, with a lagoon area of about 461 km2. The whole atoll is surrounded by an almost continuous coral reef. It was formed by a volcanic island emerging over the hot spot. Over time, the very heavy volcanic high island subsided under its own weight and was eroded by wind and rain, while the coral reef that surrounds it has grown due to the continual formation of new coral. Approximately 0.5 million years elapsed between the birth of the volcano and the atoll stage, but atoll growth can keep up with subsidence for tens of millions of years (50 million years for atolls of the western Tuamotu). The volcanoes that form the base of the Tikehau atoll reef were probably formed 80-85 million years ago, and their activity ceased in the Late Cretaceous to Early Eocene (70-50 million years ago; Rougerie *et al.,* 1997).

Plot environmental variables.

Altitude was measured in each plot using a hand-held GPS. We used estimates of mean annual precipitation (MAP) and mean annual temperature (MAT) from a set of global station measures from 1961-1990 interpolated onto a 0.5 degree grid (Mitchell and Jones 2005). Analysis of soil nutrient content from topsoil samples followed International Organization for Standardization (ISO) protocols. Total nitrogen (N) content (%) was determined based on ISO 13878: 1998. The Mehlich III procedure was used to determine soil available phosphorus (P): P was extracted by reaction with acetic acid and fluoride compounds. Chemical analyses were performed at the laboratory of the Agricultural Research Centre in Saku, Harjumaa,

Estonia. For Tasmanian plots, where soil samples were not available, we used model estimates of pH, P and N (Shangguan *et al.,* 2014). We used data from Davison *et al.* (2015) that had been collected and analysed in the same manner in order to compile equivalent soil chemistry data for mainland plots sampled in that study; and used data from Shangguan *et al.* (2014) to estimate soil parameters in mainland plots where data were missing (18 % – 25 % plots depending on the parameter).

Molecular methods

We used a metabarcoding approach to identify different organisms occurring in individual samples. Current technology limits the length of marker regions that can be amplified and sequenced, meaning that sequencing depth to some extent comes at the expense of deep taxonomic resolution. The approach used for molecular analysis exactly matches that employed in Davison *et al.* (2015). DNA was extracted from 70 mg of dried roots from each plant individual with PowerSoil-htp™ 96 Well Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) with the following modifications used by Davison *et al.,* 2015: roots were milled to powder with 3 mm tungsten carbide beads on a Mixer Mill MM400 (Retsch GmbH, Haan, Germany); to increase DNA yield, bead plates were shaken at a higher temperature (60° C) than in the default protocol and final elution was performed twice. Glomeromycota nuclear SSU rRNA gene sequences were amplified from root DNA extracts using the primers NS31 and AML2 (Simon *et al.*, 1992; Lee *et al.*, 2008). We used a set of 8-base-pair bar-codes designed following Parameswaran *et al.* (2007), where each barcode in the set differed from all others in at least two nucleotide positions. A two-step PCR procedure was conducted: in the first PCR reaction PCR primers were linked to bar-codes and partial 454-sequencing adaptors A and B; in the second reaction the full 454-adaptors A and B served as PCR primers, completing the full 454-adaptor+bar-code+PCR primer construct. Thus, the composite forward primer in the first PCR reaction was: 5’ GTCTCCGACTCAG (NNNNNNNN)*TTGGAGGGCAAGTCTGGTGCC* 3’; and the reverse primer: 5’ TTGGCAGTCTCAG (NNNNNNNN)*GAACCCAAACACTTTGGTTTCC* 3’, where the A and B adaptors are underlined, the bar-code is indicated by N-s in parentheses, and the specific primers NS31 and AML2 are shown in italics. The 10 x diluted product of the first PCR reaction was used in the second PCR with 454-adaptors A (5'-

CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and B (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') serving as PCR primers. The PCR reactions were performed in a total volume of 10 µl containing 5 µl of HotStarTaq Master Mix (Qiagen Gmbh, Germany), 0.2 µM each of the primers and 1 µl of template DNA. The reactions were run on a 2720 Thermal Cycler (Applied Biosystems, USA) under the following conditions: 95°C for 15 min; five cycles of 42°C for 30 s, 72°C for 90 s, 92°C for 45 s; 35 (first PCR) or 20 (second PCR) cycles of 65°C for 30 s, 72°C for 90 s, 92°C for 45 s; followed by 65°C for 30 s and 72°C for 10 min. PCR products were separated by electrophoresis through a 1.5% agarose gel in 0.5 × TBE, and the PCR products were purified from the gel using the QIAquick Gel Extraction kit (Qiagen Gmbh, Germany) and further purified with Agencourt® AMPure® XP PCR purification system (Agencourt Bioscience Co., Beverly, MA, USA). DNA was quantified using NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). A total of 250 ng of the resulting DNA mix was sequenced on a Genome Sequencer FLX System, using Titanium Series reagents (Roche Applied Science) at GATC Biotech (Constanz, Germany).

Bioinformatics

Bioinformatics followed the approach described in Davison *et al.* (2015). Sequencing reads were retained only if they carried the correct barcode and forward primer sequence and were ≥ 199 bp long (including the barcode and primer sequence). 1,044,180 reads met these minimum quality criteria. All reads were trimmed to 520 bp in order to remove the reverse primer sequence where it was present. 17,011 potential chimeras were identified using UCHIME (v7.0.1090, Edgar *et al.,* 2011.) in reference database mode with the default settings, and excluded from the analyses.

We then used an open-reference operational taxonomic unit (OTU) picking approach to match obtained reads against taxa in the Maarj*AM* database of published Glomeromycotina SSU rRNA gene sequences (http://maarjam.botany.ut.ee; Öpik *et al.,* 2010). The Maarj*AM* database contains representative SSU rRNA gene (and other marker regions) sequences from environmental DNA and morphologically described taxa. As of October 2015 it contained 6,064 SSU rRNA gene sequences corresponding to 352 SSU sequence-based taxa,

or so-called virtual taxa (VT), as well as 73 further phylogroups not yet assigned to VT. Sequencing reads were assigned to phylogroups or VT (hereafter VT) by conducting a BLAST (BLAST+ v2.5.0, Camacho *et al.,* 2009) search (soft masking of DUST filter) against the Maarj*AM* database with the following criteria required for a match: sequence similarity ≥ 97%; an alignment length not differing from the length of the shorter of the query (454-read) and subject (reference database sequence) sequences by more than 5%; and a BLAST e-value < 1e-50. VT are phylogenetically-defined groupings at approximately the species level or slightly above. Sequence data from isolates of known identity have been shown to represent similar numbers of VT and morphospecies (Öpik and Davison, 2016). However there is not a one to one correspondence, and VT are known to merge closely related morphospecies in some lineages (Öpik *et al.,* 2013; Öpik and Davison, 2016). Nonetheless, the rank of VT has been shown to capture ecologically-relevant responses to environmental gradients (Powell *et al.,* 2011).

We searched for previously undescribed taxa among the reads not receiving a match against the Maarj*AM* database. First, such unmatched reads were clustered at 99% with 90% length coverage using BLASTclust 2.2.26 from the legacy NCBI BLAST tools (Altschul *et al.,* 1990); then the longest sequence from each cluster was submitted to a BLAST search against the non-redundant INSDC. Clusters were considered putative Glomeromycotina if they gained a best hit against a Glomeromycotina sequence with similarity ≥ 90%, an alignment length not differing from the length of the shorter of the query (454-read) and subject (reference database sequence) sequences by more than 10%; and a BLAST e-value < 1e-50. New VT were identified among the putative Glomeromycotina clusters following the procedure used to update the Maarj*AM* database, as described in Öpik *et al.* (2010); briefly, by aligning (muscle v3.6; Edgar 2004) the four longest sequences from each cluster with sequences from the Maarj*AM* database; constructing Neighbor-joining (NJ) phylogenies (F84 model with gamma substitution rates) and identifying those clusters that were divergent at ≥ 97 % sequence similarity in a well-supported phylogenetic analysis. This approach identified 6 previously undescribed VT: 3 Glomeraceae; 1 Claroideoglomeraceae; 1 Archaeosporaceae; 1 Paraglomeraceae.

A final BLAST search was carried out on a read set containing (i) the quality-controlled reads from this analysis of island samples and (ii) the quality controlled reads from Davison *et al.*

(2015), which comprised a global set of samples from mainland locations. The parameters of this BLAST matched those described above, except that the reference database consisted of the Maarj*AM* database supplemented with 4 sequences from each previously undescribed VT. 1,695,324 reads were matched against virtual taxa (VT) in this supplemented Maarj*AM* database (island: 681,992; mainland: 1,013,332). Samples with < 20 reads and VT represented by a single read in the entire data set were removed. The resulting data matrices were used for analysis: 419 island samples comprising 681,733 reads; and 831 mainland samples from Davison *et al.* (2015), comprising 1,012,742 reads (a combined table including island and mainland sequencing data is provided in Table S2).

Phylogenetic analysis

We reconstructed a Bayesian phylogeny of AM fungi using BEAST (v1.8.0; Drummond *et al.,* 2012). The analysis incorporated existing AM fungal VT type sequences and a representative sequence from each of the 6 previously unrecorded taxa discovered in this analysis (the longest available sequence for each). Sequences were aligned using MAFFT (v7.245, Katoh and Standley, 2013) and subjected to phylogenetic analysis using the GTR + I + G nucleotide substitution model (selected on the basis of BIC; jmodeltest; Posada *et al.,* 2008). The posterior simulation-based analogue of Akaike's information criterion (AICM; Raftery *et al.,* 2007) was used to select a log-normal relaxed clock model with a coalescent tree model (log-normal relaxed clock). Final MCMC analysis consisted of nine separate chains of 20,000,000 steps, with approximately 25% of steps removed as burn-in from each chain. The results were summarised on a maximum clade credibility tree.

To estimate the timing of evolutionary events we used the fossil calibration proposed by Davison *et al.,* (2015): a Gaussian prior (mean = 505 Mya, standard deviation = 54 million years) for the age of the root, such that 95% of the probability density falls between 400 Mya and 610 Mya. These dates respectively represent the ages of fossils resembling multiple extant AM fungal lineages (Dotzler *et al.*, 2009) and an estimate for the timing of divergence between Glomeromycotina and other fungal lineages (Redecker *et al.,* 2000). We found that, on average, individual AM fungal VT diverged from their sister (most closely related) VT in the last 20 million years (mean = 17.8 Mya; interquartile range = 8.6 – 19.4 Mya). This time

scale appears approximately comparable with the estimated timing of divergences between sister plant species (mean = 15.9 Mya; interquartile range = 3.6 – 15.7 Mya; n = 5122 species) or the most recent common ancestor of plant genera (mean = 19.1 Mya; interquartile range = 6.7 – 22.2; n = 514 genera) in the European flora (Durka and Michalski, 2012).

Virtual taxon life history characteristics

We used the Maarj*AM* database to identify those VT that could be associated with an AM fungal morphospecies (i.e. species described on the basis of spore traits from a cultured organism). VT were defined as 'cultured' if they contained sequences derived from an isolate of known morphospecies identity. While morphospecies clearly exhibit a variety of life history strategies (Johnson *et al.,* 1991), abundantly sporulating taxa are assumed to be efficient dispersers, and cultured VT have been associated with anthropogenic habitats (Ohsowski *et al.,* 2014). We identified 60 cultured VT in the Maarj*AM* database, of which 46 were recorded in island or mainland samples, contributing 25% of reads in the final dataset. The cultured trait was moderately conserved (D statistic for phylogenetic signal in binary traits= 0.71; Fritz and Purvis, 2010), and cultured VT appeared to be overrepresented in the families Gigasporaceae, Diversisporaceae and Acaulosporaceae. We compiled existing data on AM fungal morphospecies spore diameter from several published sources. Since available sources to some degree update and duplicate one another, we checked for information about AM fungal morphospecies (mean) spore diameter in the following sequence: culture collection INVAM (http://invam.wvu.edu/), web page of Janusz Blaszkowski (www.zor.zut.edu.pl/Glomeromycota), an unpublished summary of species descriptions prepared by Gisela Cuenca in 2002 (see entries in Table S3), Schenck and Perez (1990). We were able to retrieve spore diameter estimates from 190 AM fungal morphospecies (Table S3). We identified 48 VT in the Maarj*AM* database that included one or more of the morphospecies with a spore diameter estimate. Of these, 27 VT could be assigned spore diameter estimates from a single morphospecies. For the remaining 21 VT, spore diameter was estimated as the mean of values taken from multiple component morphospecies (intraclass correlation coefficient = 0.52). Thirty-seven of these VT were recorded in island or mainland samples, contributing 22 % of reads in the final dataset. The spore diameter trait

was moderately conserved among VT for which a diameter estimate was available (Blomberg's K = 0.19; Blomberg *et al.,* 2003), with Gigasporaceae VT consistently exhibiting the largest spores.

Statistical modelling and randomisation

Modelling interpretation: By collecting samples from point locations on each island, we aimed to estimate characteristics of local AM fungal communities. While additionally estimating characteristics of the pools of AM fungi inhabiting entire islands (especially gamma diversity) would have been desirable it would have been extremely challenging, given the size of study islands and the complexity of measuring soil communities. In the context of island biogeography, a point sampling approach minimises the effect of heterogeneity on diversity (larger islands are expected to be more heterogeneous and present more potential niches). Consequently the possible effects of isolation and area *per se* should be more evident.

PERMANOVA: statistical significance was calculated using randomisation to account for spatial dependence between samples. In each random iteration (n=999) entire plots were randomly reassigned to the island or mainland set, and P values (P_{rand}) were estimated from the number of randomised pseudo-F statistics more extreme than observed value.

Distance decay: since each mainland site contributed to multiple data points (including some that were paired with multiple islands; Fig S1) we calculated randomised P values to assess statistical significance. For the geographic distance analyses, in each random iteration, the spatial coordinates of sites were shuffled within each continent, so that mainland sites were represented the same number of times in the raw data and in each random iteration (n=999). For the island area analyses, the paired mainland site used for comparison was shuffled among islands. Randomised F statistics associated with geographic distance or area (and in the case of the distance analysis a variable distinguishing island-mainland and mainland-mainland pairs) were then calculated from linear models in each random iteration. We estimated P values (P_{rand}) from the number of randomised F statistics more extreme than the corresponding observed values. We conducted IBE analyses by first calculating a distance-based redundancy analysis (dbRDA, function capscale from vegan; using Sørensen or Bray-Curtis distance) against the environmental variable set. We then calculated mean

pairwise distances among the residuals from this model; and these distances were used as dependent variables in the geographic distance and area analyses.

LMMs and GLMMs: We calculated raw richness as well as the asymptotic estimators for exponential Shannon diversity and reciprocal Simpson diversity (i.e., Hill numbers 1-2), using the iNEXT package in R (Chao *et al.,* 2014). Raw richness estimates approached estimates of asymptotic richness (Chao; raw richness was on average 85 % of Chao richness) and were highly correlated with the asymptotic estimator ($r = 0.93$, $P < 0.001$). However, extrapolation using Chao can be very imprecise beyond double the sample size (Chao *et al.,* 2014; and this might affect some of our lowest sample sizes), so we do not present Chao in the manuscript.

Site and plot were included as nested random effects in all models, while island identity was included as an additional nested random effect in models containing only island samples. Generalised linear mixed models with a binomial error structure were used to model proportional response variables (i.e. the proportions of Glomerales or cultured VT). Models were constructed using function *glmer* in R package lme4 (Bates *et al.,* 2015), except for the model examining mean spore diameter. For this model, we used function *lme* from R package nlme (Pinheiro *et al.,* 2016) and constructed a heterogeneous variance model at the level of plots using the varIdent argument. The significance of model terms was calculated after accounting for all other model terms, with degrees of freedom for linear mixed models estimated using the Kenward Roger approximation (except for the lme models where in-out rules produced similar degrees of freedom). Island area and remoteness were log transformed while environmental and climatic variables were scaled prior to inclusion in models.

In addition to island remoteness and area, island origin (continental vs oceanic) is a potential driver of diversity patterns (Whittaker and Fernandez-Palacios, 2007). However, since island origin was correlated with island remoteness (islands of continental origin [mean = 346 km] tended to be less remote than islands of oceanic origin [mean = 2675 km]) we did not include this variable in analyses. Nonetheless, we distinguish islands of continental and oceanic origin on figures.

Supplementary references

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Binard N, Maury RC, Guille G, Talandier J, Gillot PY, Cotten J. (1993). Mehetia Island, South Pacific: geology and petrology of the emerged part of the Society hot spot. *J Volcanol Geotherm Res* **55**: 239-260.

Blais S, Guille G, Maury RC, Guillou H, Miau D, Cotten J. (1997). Geology and petrology of Raiatea Island (Society islands, French Polynesia). *C R Acad Sci Paris Sér* **324**: 435-442.

Blais S, Guille G, Guillou H, Chauvel C, Maury RC, Pernet G. *et al.,* (2002). The island of Maupiti: the oldest emergent volcano in the Society hot spot chain (French Polynesia). *Bull Soc géol France* **173**: 45-55.

Blomberg SP, Garland Jr T, Ives AR. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717-745

Bover P, Quintana J, Alcover JA. (2008). Three islands, three worlds: paleogeography and evolution of the vertebrate fauna from the Balearic Islands. *Quat Int* **182**: 135-144.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K. *et al.,* (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell, RK. *et al.,* (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol Monograph* **84**: 45-67.

Dallmann WK. (ed.) (2015). *Geoscience Atlas of Svalbard*. Norwegian Polar Institute, Report Series 148, 1-292.

Dotzler N, Walker C, Krings M, Hass H, Kerp H, Taylor TN. *et al.,* (2009). Acaulosporoid glomeromycotan spores with a germination shield from the 400-million-year-old Rhynie chert. *Mycol Prog* **8**: 9-18.

Drummond, AJ, Suchard MA, Xie D, Rambaut A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**: 1969-1973

Durka W, Michalski SG. (2012). Daphne: a dated phylogeny of a large European flora for phylogenetically informed ecological analyses. *Ecology* **93**: 2297-2297.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194-2200

Fritz SA, Purvis A. (2010). Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv Biol* **24**: 1042-1051.

Grandcolas P, Murienne J, Robillard T, Desutter-Grandcolas L, Jourdan H, Guilbert E. *et al.,* (2008). New Caledonia: a very old Darwinian island? *Philos Trans R Soc Lond* B *Biol Sci* **363**: 3309–3317.

Hildenbrand A, Gillot P-Y, Le Roy I. (2004). Volcano-tectonic and geochemical evolution of an oceanic intra-plate volcano: Tahiti-Nui (French Polynesia). *Earth Planet Sci Lett* **217**: 349–365.

Hill RS. (1990). Sixty million years of change in Tasmania's climate and vegetation. *Tasforests* **2**: 89-98.

Johnson NC, Zak DR, Tilman D, Pfleger FL. (1991). Dynamics of vesicular-arbuscular mycorrhizae during old field succession. *Oecologia* **86**: 349-358.

Katoh K, Standley DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

Lee J, Lee S, Young JPW. (2008). Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiol Ecol* **65**: 339-349.

Lowry II PP. (1998). Diversity, endemism, and extinction in the flora of New Caledonia: a review. In: Peng, C-I., Lowry II PP. (eds). *Rare, threatened, and endangered floras of Asia and the Pacific rim, vol. 16*, 181–206. Taipei: Institute of Botany, Academia Sinica.

Mitchell TD, Jones PD. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* **25**: 693-712.

Münch P, Lebrun J-F, Cornée J-J, Thinon I, Guennoc P, Marcaillou B. *et al.,* (2013). Pliocene to Pleistocene carbonate systems of the Guadeloupe archipelago, French Lesser Antilles: a land and sea study. *Bull Soc géol Fr* **184**: 99–110.

Neall VE, Trewick SA. (2008). The age and origin of the Pacific islands: a geological overview. *Philos Trans Roy Soc B – Biol Sci* **363**: 3293–3308.

Öpik M, Davison J, Moora M, Zobel M. (2013). DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. *Botany* **92**: 135-147.

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M. *et al.,* (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* **35**

Posada D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253-1256.

Raftery AE, Newton MA, Satagopan JM, Krivitsky PN. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. in: *Bayesian statistics 8* (eds Bernardo JM. *et al*.), 1-45, Oxford University Press, Oxford, UK.

Ramalho RAS. (2011). *Building the Cape Verde Islands*. Springer, Berlin.

Redecker D, Kodner R, Graham LE. (2000). Glomalean fungi from the Ordovician. *Science* **289**: 1920-1921.

Rosentau A, Vassiljev J, Hang T, Saarse L, Kalm V. (2009). Development of the Baltic Ice Lake in eastern Baltic. *Quat Int* **206**: 16-23.

Rougerie F, Fichez R, Dejardin P. (1997). Geomorphology and hydrogeology of selected islands of French Polynesia: Tikehau (atoll) and Tahiti (barrier reef). In: Vacher HL, Quinn T. (eds). *Geology and Hydrogeology of Carbonate Islands. Developments in Sedimentology 54*, 475-502. Elsevier, Amsterdam.

Sæmundsson K. (1979). Outline of the geology of Iceland. *Jökull* **29**: 7-28.

Schenck NC, Perez Y. (1990). *Manual for the identification of VA mycorrhizal fungi*. Synergistic Publications, FL, USA

Shangguan W, Dai Y, Duan Q, Liu B, Yuan H. (2014). A Global Soil Data Set for Earth System Modeling. *J Adv Model Earth Sy* **6**: 249-263.

Simon L, Lalonde M, Bruns TD. (1992). Specific amplification of 18S fungal ribosomal genes from VA endomycorrhizal fungi colonizing roots. *Appl Environ Microbiol* **58**: 291-295.

Van Der Geer AAE, Dermitzakis M, De Vos J. (2006). Crete before the Cretans: the reign of dwarfs. *Pharos* **13**: 121-132.

Vassiljev J, Saarse L. (2013). Timing of the Baltic Ice Lake in the eastern Baltic. *Bull Geol Soc Finland* **85**: 9-18.

Other Supplementary Information

Figure S1. Maps of island sites and paired mainland sites. Each island was paired with the studied mainland site that was geographically closest (see Fig 1).

Figure S2. NMDS based on Sørensen distance between AM fungal communities associating with plant species at island and mainland sites. For each island, samples from the focal island (square), the paired mainland site (triangle) and continent (dark shaded ellipse) are highlighted; other mainland samples (small points) and continents (light shaded ellipses) are also shown (Africa = blue, Asia = red, Europe = black, North America= yellow, Oceania = brown, South America = green).

Figure S3. Bayesian phylogeny of Glomeromycotina virtual taxa (VT; SSU rRNA gene sequences). The relative abundance of reads derived from each VT is shown for each island and mainland site (i.e. symbol areas sum to 1 within each column). Sites are coloured according to their continent/region of origin; with island sites shown in a darker shade.

Figure S4. Abundance-weighted analyses. Mean compositional distance (Bray-Curtis) between plant-root associated AM fungal communities as a function of the geographic distance separating them (a-b). Points represent community pairs: islands and paired mainland sites (large; blue or green) or pairs of mainland sites within the same continent (small grey points). a) island biogeography (IB) model not accounting for environmental characteristics; b) island biogeography and environment (IBE) model accounting for environmental characteristics. Separate regression lines are shown for island-mainland (solid) and mainland-mainland (dashed) points. Compositional distances in the IBE model were calculated among the residuals of a dbRDA model against measured environmental variables. Relationships between island remoteness and several abundance-weighted taxonomic and life history characteristics of plant-root associated AM fungal communities (c-f): c) the mean proportion of Glomerales VT, d) the mean pairwise phylogenetic distance (MPD) between VT, e) the proportion of cultured VT, and f) the mean spore diameter of VT. Lines show predicted values from island biogeography (IB) generalised linear mixed models. Blue symbols indicate oceanic islands; green symbols indicate continental islands.

Figure S5. Relationship between island area and several taxonomic and life history characteristics of plant-root associated AM fungal communities (not weighted by organism abundance): a, e) the mean proportion of Glomerales VT, b, f) the mean pairwise phylogenetic distance (MPD) between VT, c, g) the proportion of cultured VT, and d, h) the mean spore diameter of VT. a-d show characteristics that are not weighted by organism abundance; e-h show abundance weighted measures. Lines show predicted values from island biogeography (IB) generalised linear mixed models. Blue symbols indicate oceanic islands; green symbols indicate continental islands.

Table S1. Sampled Plots. Sampled plant species and measured environmental variables are reported for each plot. Numbers in parentheses after each plant species indicate the number of samples used in analysis (i.e. meeting quality control criteria). VT shows the number of AM fungal virtual taxa recorded in each plot. [In a separate spreadsheet].

Table S2. Sequencing data used for analysis. The data table shows read counts of different VT in samples after quality control procedures, BLAST searches against VT in the Maarj*AM* database (Öpik *et al.,* 2010) and identification of previously unknown VT. Taxa in bold are

new VT; Pacific samples in bold are those retained in analysis of richness (see methods for details). [In a separate spreadsheet].

Table S3. Data on AM fungal morphospecies spore characteristics and VT containing a described morphospecies. [In a separate spreadsheet].

Table S4. The effect of island vs mainland location and island characteristics (area and remoteness) on parameters of AM fungal communities worldwide based on relative abundance data. The island biogeography model (IB model) tests the effect of island biogeographic predictors (island vs mainland location (Set); or island area and remoteness) in isolation; while the island biogeography and environment model (IBE model) tests their effects after accounting for the effect of environmental and climatic variables. Fixed effect coefficients (b), test statistics (F with Kenward-Roger estimated degrees of freedom or χ^2) and statistical significance from linear or generalised linear mixed models are presented in the table. ** $P < 0.01$ * $P < 0.05$. $P < 0.1$ ($P < 0.05$ in bold). The effect of each variable is tested after accounting for all other variables (Type II). MAT – mean annual temperature; MAP – mean annual precipitation. [In a separate spreadsheet].

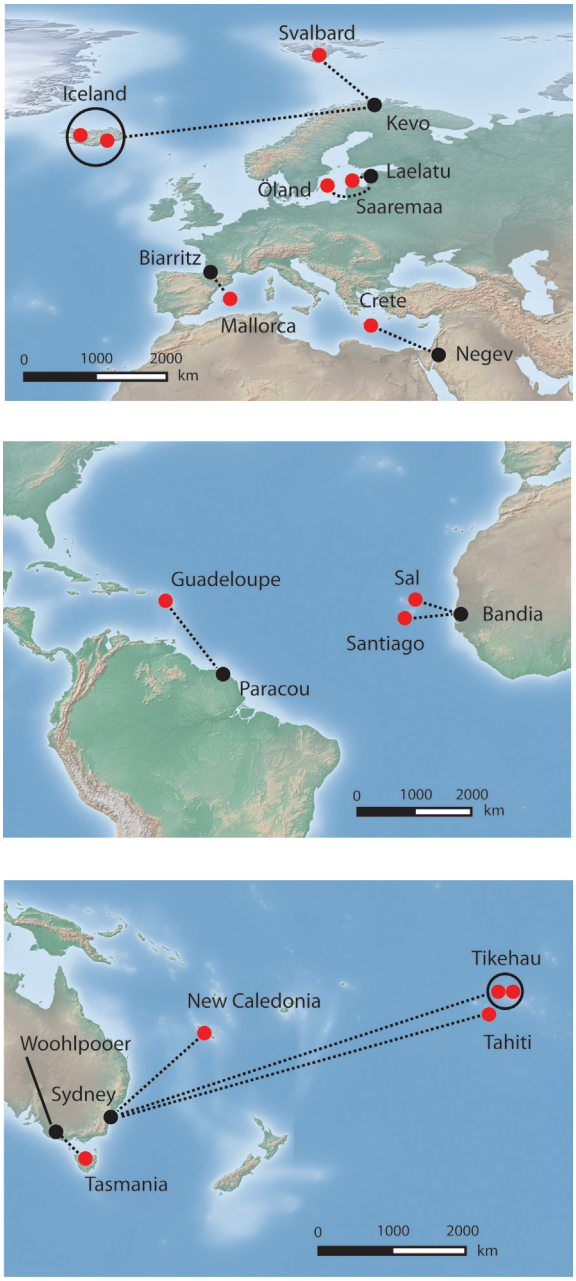


Figure S1. Maps of island sites and paired mainland sites. Each island was paired with the studied mainland site that was geographically closest (see Fig 1).

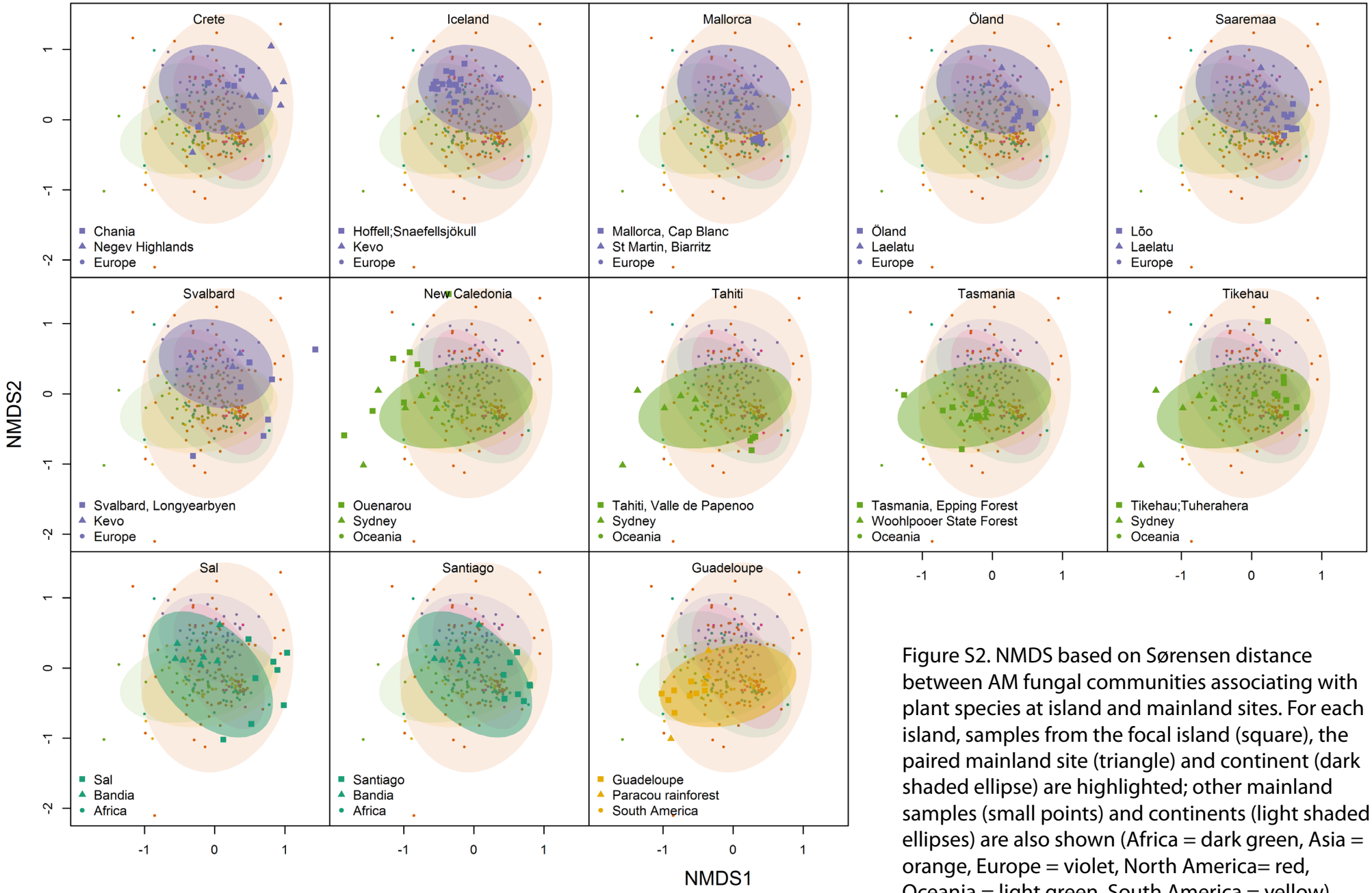
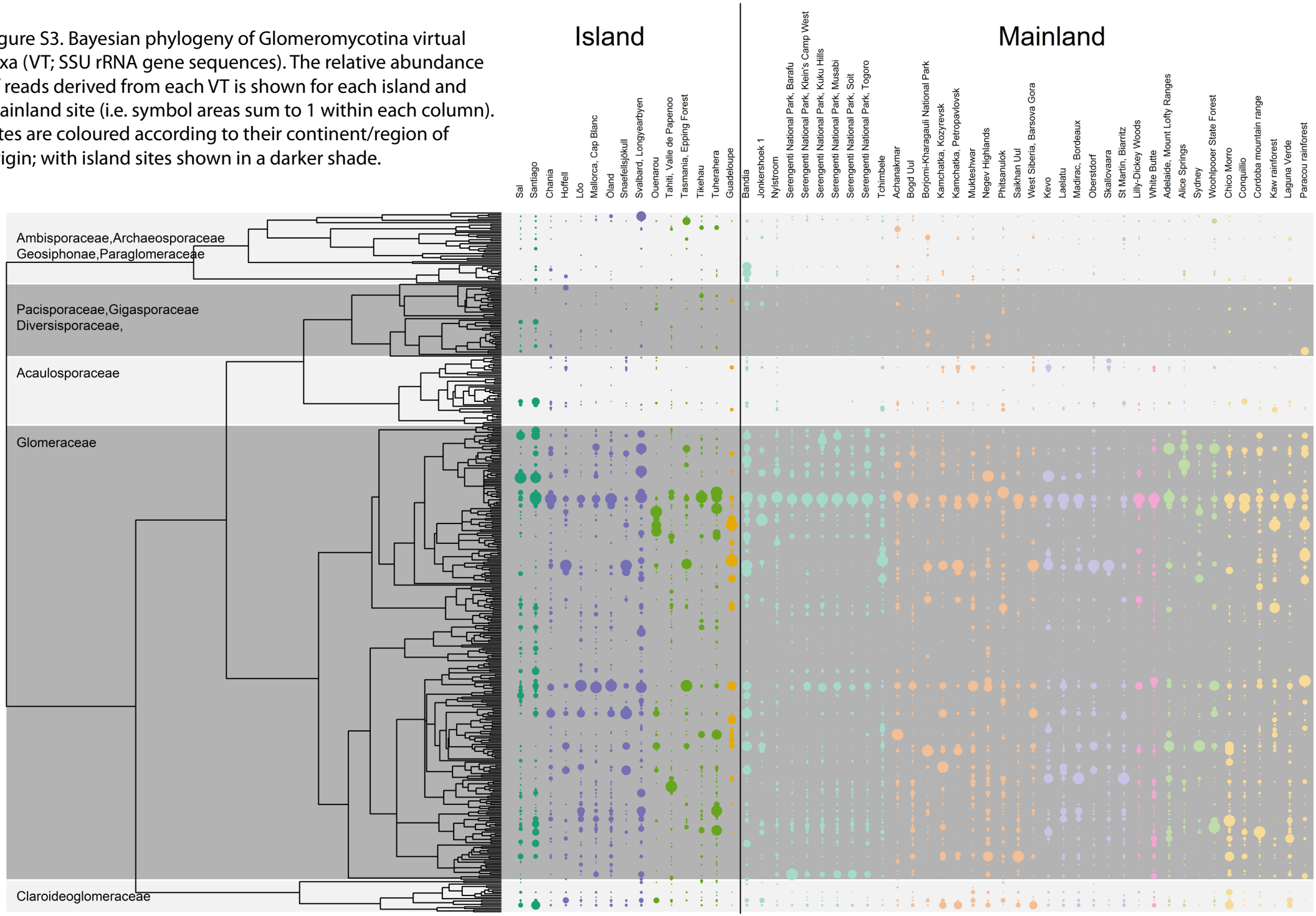


Figure S2. NMDS based on Sørensen distance between AM fungal communities associating with plant species at island and mainland sites. For each island, samples from the focal island (square), the paired mainland site (triangle) and continent (dark shaded ellipse) are highlighted; other mainland samples (small points) and continents (light shaded ellipses) are also shown (Africa = dark green, Asia = orange, Europe = violet, North America= red, Oceania = light green, South America = yellow).

Figure S3. Bayesian phylogeny of Glomeromycotina virtual taxa (VT; SSU rRNA gene sequences). The relative abundance of reads derived from each VT is shown for each island and mainland site (i.e. symbol areas sum to 1 within each column). Sites are coloured according to their continent/region of origin; with island sites shown in a darker shade.



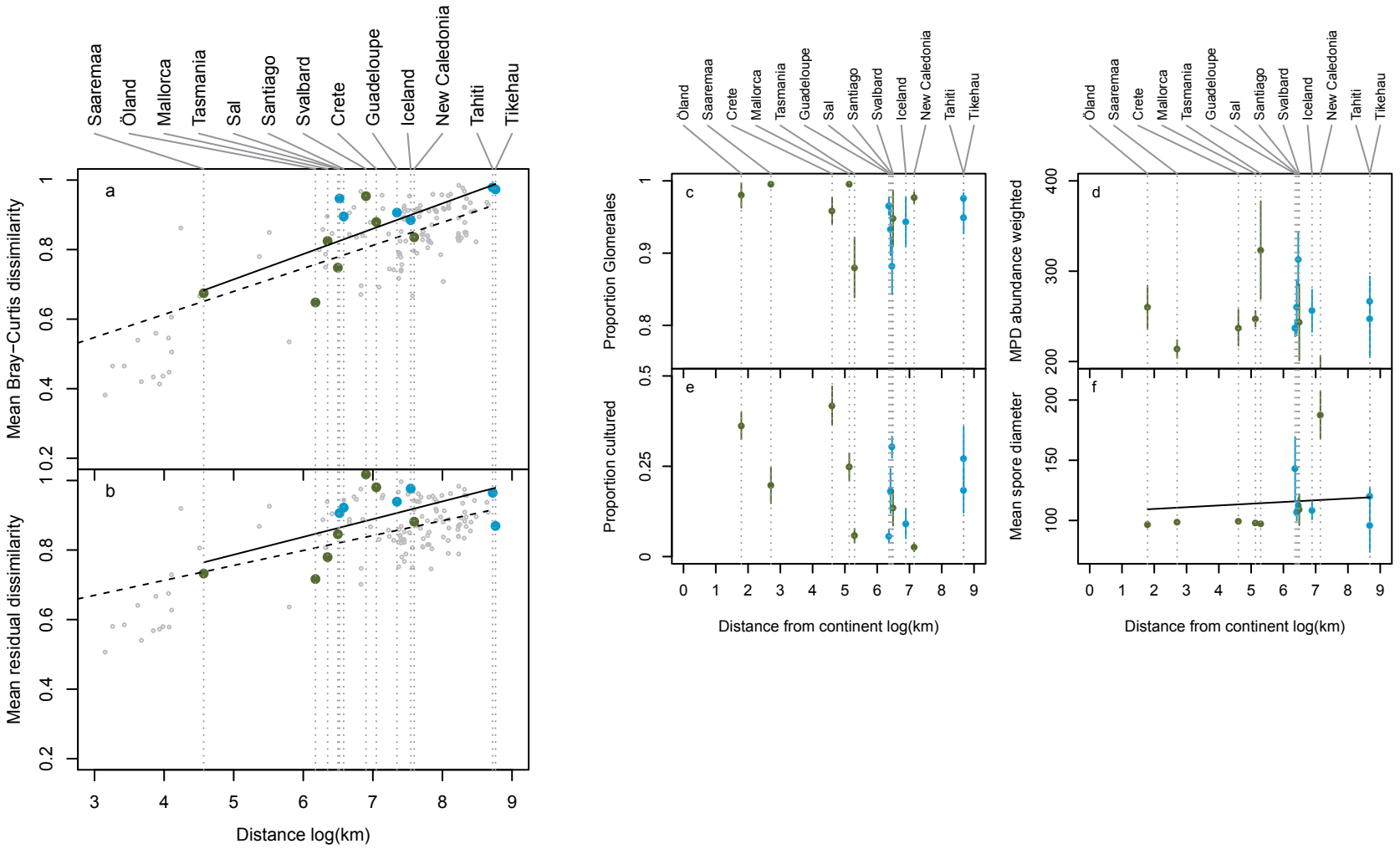


Figure S4. Abundance-weighted analyses. Mean compositional distance (Bray-Curtis) between plant-root associated AM fungal communities as a function of the geographic distance separating them (a-b). Points represent community pairs: islands and paired mainland sites (large; blue or green) or pairs of mainland sites within the same continent (small grey points). a) island biogeography (IB) model not accounting for environmental characteristics; b) island biogeography and environment (IBE) model accounting for environmental characteristics. Separate regression lines are shown for island-mainland (solid) and mainland-mainland (dashed) points. Compositional distances in the IBE model were calculated among the residuals of a dbRDA model against measured environmental variables. Relationships between island remoteness and several abundance-weighted taxonomic and life history characteristics of plant-root associated AM fungal communities (c-f): c) the mean proportion of Glomerales VT, d) the mean pairwise phylogenetic distance (MPD) between VT, e) the proportion of cultured VT, and f) the mean spore diameter of VT. Lines show predicted values from island biogeography (IB) generalised linear mixed models. Blue symbols indicate oceanic islands; green symbols indicate continental islands.

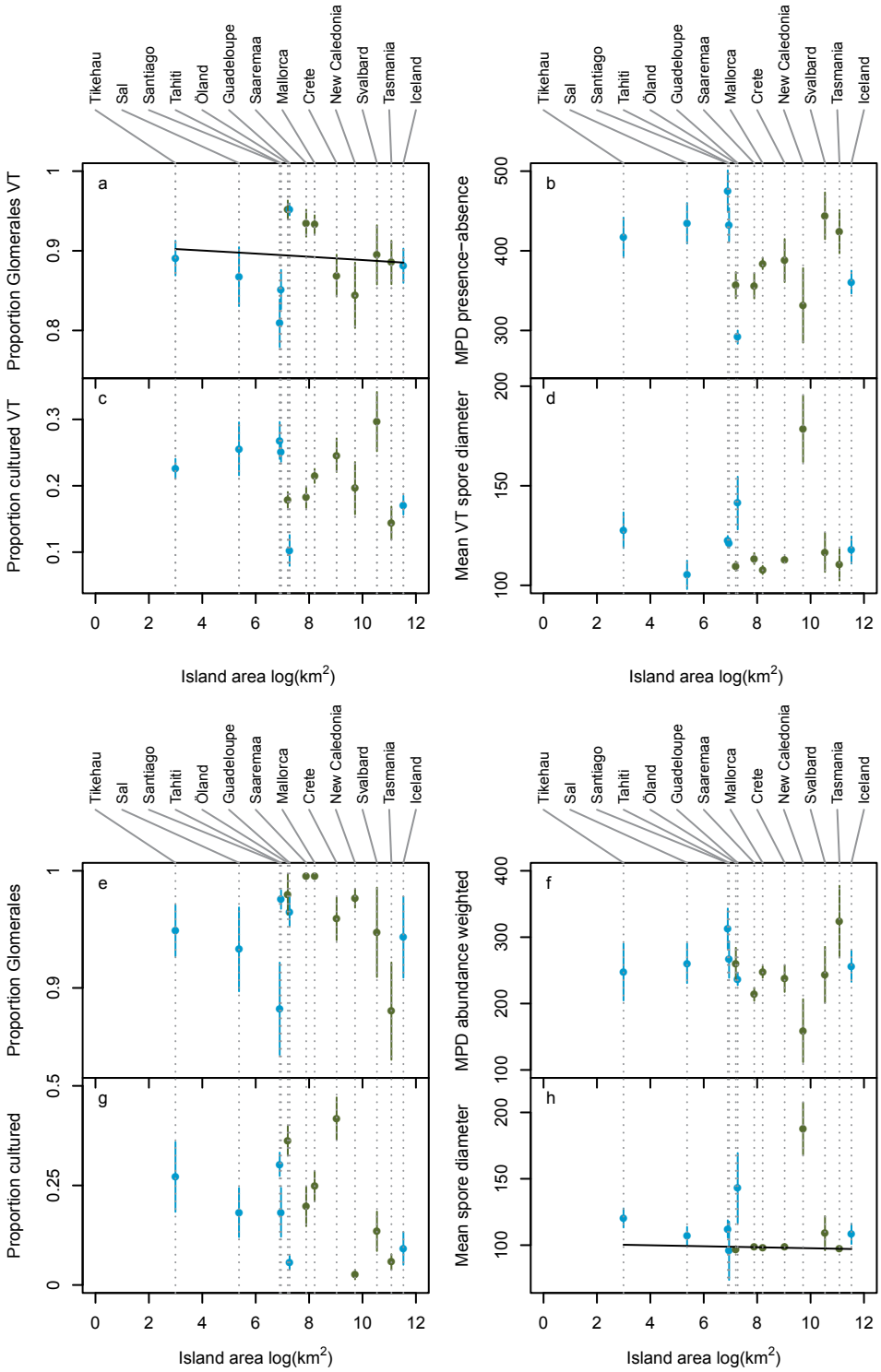


Figure S5. Relationship between island area and several taxonomic and life history characteristics of plant-root associated AM fungal communities: a, e) the mean proportion of Glomerales VT, b, f) the mean pairwise phylogenetic distance (MPD) between VT, c, g) the proportion of cultured VT, and d, h) the mean spore diameter of VT. a-d show characteristics that are not weighted by organism abundance; e-h show abundance weighted measures. Lines show predicted values from island biogeography (IB) generalised linear mixed models. Blue symbols indicate oceanic islands; green symbols indicate continental islands.